



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Quantitative Shotgun Proteomics Unveils Candidate Novel Esophageal Adenocarcinoma (EAC)-specific Proteins

Citation for published version:

O'Neill, JR, Pak, H, Pairo-castineira, E, Save, V, Paterson-brown, S, Nenutil, R, Vojtšek, B, Overton, I, Scherl, A & Hupp, TR 2017, 'Quantitative Shotgun Proteomics Unveils Candidate Novel Esophageal Adenocarcinoma (EAC)-specific Proteins', *Molecular & Cellular Proteomics (MCP)*, vol. 16, no. 6, pp. 1138-1150. <https://doi.org/10.1074/mcp.M116.065078>

Digital Object Identifier (DOI):

[10.1074/mcp.M116.065078](https://doi.org/10.1074/mcp.M116.065078)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Molecular & Cellular Proteomics (MCP)

Publisher Rights Statement:

Deposit permitted by the publisher.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Title: Quantitative Shotgun Proteomics Unveils Candidate Novel Oesophageal Adenocarcinoma-Specific Proteins

Authors: J Robert O'Neill^{1,2}, Hui-Song Pak³, Erola Pairo-Castineira^{4,5}, Vicki Save⁶, Simon Paterson-Brown², Rudolf Nenutil⁷, Bořivoj Vojtěšek⁷, Ian Overton^{4,5}, Alex Scherl³, Ted R Hupp^{1,7}

Affiliations: ¹ Edinburgh Cancer Research Centre at the Institute of Genetics and Molecular Medicine, Edinburgh University. ² Department of Surgery, Royal Infirmary of Edinburgh. ³ Department of Human Protein Sciences, Faculty of Medicine, University of Geneva. ⁴ Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh. ⁵ MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh University. ⁶ Department of Pathology, Royal Infirmary of Edinburgh. ⁷ Regional Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, Brno.

Correspondence: Mr. J Robert O'Neill, Clinical Lecturer, Department of Surgery, Edinburgh Royal Infirmary, 51 Little France Crescent, Edinburgh EH16 4SA. Telephone: +44 131 537 1000.

Email: Robert.o'neill@ed.ac.uk

Running Title: Identifying oesophageal adenocarcinoma-specific proteins

Abbreviations: Oesophageal adenocarcinoma (OAC), oesophagogastric junction (OGJ), computed tomography (CT), positron emission tomography (PET), endoscopic ultrasound (EUS), Tandem Mass Tag (TMT), immunohistochemistry (IHC), room temperature (RT), filter-aided sample preparation (FASP), formic acid (FA), distilled water (dH₂O), nominal molecular weight cut-off (NMWCO), ultra-performance liquid chromatography (UPLC), ratio of OAC to normal oesophagus expression (TvO), ratio of OAC to normal gastric expression (TvG), oesophageal squamous cell carcinoma (OSCC), lymphovascular invasion (LVI), perineural invasion (PNI)

SUMMARY

Oesophageal cancer is the eighth most common cancer worldwide and the majority of patients have systemic disease at presentation. Oesophageal adenocarcinoma (OAC), the predominant subtype in western countries, is largely resistant to current chemotherapy regimens. Selective markers are needed to enhance clinical staging and to allow targeted therapies yet there are minimal proteomic data on this cancer type.

After histological review, lysates from OAC and matched normal oesophageal and gastric samples from seven patients were subjected to LC MS/MS after tandem mass tag labelling and OFFGEL fractionation. Patient matched samples of OAC, normal oesophagus, normal stomach, lymph node metastases and uninvolved lymph nodes were used from an additional 115 patients for verification of expression by immunohistochemistry (IHC).

Over six thousand proteins were identified and quantified across samples. Quantitative reproducibility was excellent between technical replicates and a moderate correlation was seen across samples with the same histology. The quantitative accuracy was verified across the dynamic range for seven proteins by immunohistochemistry (IHC) on the originating tissues. Multiple novel tumour-specific candidates are proposed and EPCAM was verified by IHC.

This shotgun proteomic study of OAC used a comparative quantitative approach to reveal proteins highly expressed in specific tissue types. Novel tumour-specific proteins are proposed and EPCAM was demonstrated to be specifically over-expressed in primary tumours and lymph node metastases compared to surrounding normal tissues. This candidate and others proposed in this study could be developed as tumour-specific targets for novel clinical staging and therapeutic approaches.

INTRODUCTION

Oesophageal cancer is the sixth leading cause of cancer death worldwide ¹ and oesophageal adenocarcinoma (OAC) has become the predominant histological subtype in western countries ^{2,3}. In the UK, 95% of patients diagnosed with OAC will die from metastatic disease and the majority are resistant, at presentation, to current platinum-based chemotherapy regimens ⁴⁻⁶.

OAC is frequently associated with both lymphatic and distant metastases yet current staging modalities including computed tomography (CT), positron emission tomography (PET) and endoscopic ultrasound (EUS) are limited in both sensitivity and specificity ⁵. Surgical resection only benefits patients with localised disease and carries a 40% risk of major morbidity and 2-3% risk of perioperative mortality ^{7,8}. The development of accurate non-invasive imaging markers of OAC would enhance clinical staging by allowing the specific detection of locoregional and distant metastases, enabling treatment stratification ⁹.

The normal squamous epithelium-lined oesophagus is vulnerable to toxic insult from the oesophageal lumen. Indeed, chronic reflux of gastric acid and bile is thought to underlie the development of columnar metaplasia - "Barrett's oesophagus", the precursor lesion of OAC ¹⁰. Although the exact molecular mechanisms of Barrett's development and oesophageal carcinogenesis remain obscure, the detection and treatment of OAC at an early stage offers the prospect of long term cure with over 80% of patients undergoing surgery for stage I oesophageal cancer surviving 5 years ¹¹.

Intriguingly many of the genetic mutations present in OAC have also been demonstrated in non-dysplastic Barrett's epithelium raising the possibility that a transcriptional change such as splicing or RNA-editing, or a post-translational modification is responsible for transformation ¹². If such a biomarker could be identified this would offer the possibility of earlier diagnosis and more effective treatment.

Characterising the proteomic changes associated with OAC may also allow novel therapies to be designed. Tumour-specific proteins have been exploited as immunotherapeutic targets in other cancer types by engendering a host response to the cancer ¹³, in some cases leading to durable responses ¹⁴.

To date, no specific markers of OAC have been identified. To identify candidate proteins de novo, expression must be measured using untargeted proteomic methods.

Quantitative proteomic methods have now been applied across many cancer tissues. Most previous proteomic studies in OAC, however, have only identified a small number of dysregulated proteins limiting the comparisons that can be made between studies or with other cancers (summarised in Table 1). Only one of these previous studies employed a quantitative shotgun proteomic strategy. The authors compared pooled biopsies of OAC, normal oesophagus, gastric adenocarcinoma and normal gastric tissue and identified 972 proteins. Although no OAC-specific protein was identified, neutrophil defensin 1, an antimicrobial peptide found in neutrophil granules, was overexpressed in both cancer types relative to normal tissue ¹⁵. This may reflect the inflammatory environment associated with these cancers.

The comparisons between OAC and normal squamous epithelium in published work reveals many dysregulated proteins, some of which represent proteins associated with glandular differentiation and some associated with carcinogenesis. Glandular-associated proteins may be expressed in gastric and intestinal epithelium and may not represent tractable targets for therapy as toxicity due to intestinal epithelial damage would be expected. It is possible that including columnar epithelium-lined gastric tissue along with squamous and OAC tissue may enable the discrimination of proteins that reflect glandular differentiation from those driving carcinogenesis.

Multi-tissue proteomic profiling has been applied across mouse tissues with relative quantitation using a super-SILAC approach ¹⁶. In this study, snap-frozen biopsies from 28 tissue types were subjected to shotgun proteomics with a spike-in, heavy-labelled mixture of all tissues obtained from the SILAC mouse. By comparing the relative expression of proteins across tissues, tissue-specific expression could be highlighted. The oesophagus was not included in this profiling effort although gastrointestinal tissues with columnar epithelia showed similar expression patterns ¹⁶. This comparative approach has also been employed in a large proteomic study of 30 human tissues by label-free quantification and again tissue-specific expression patterns identified ¹⁷.

This biomarker discovery study therefore used a quantitative shotgun proteomic strategy to evaluate protein expression in OAC and adjacent matched normal squamous and gastric tissues from seven patients. By quantifying the relative expression between OAC and normal oesophagus and OAC and

normal stomach, proteins aberrantly expressed in OAC were identified. The accuracy of this approach was confirmed by immunohistochemistry for multiple candidates and a potential tumour biomarker verified in a cohort of 115 OAC patients with matched normal and metastatic tissues.

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

Fresh frozen biopsies representing macroscopically normal oesophagus, normal stomach and oesophageal adenocarcinoma tissue were prospectively collected from resection specimens from seven patients undergoing neoadjuvant chemotherapy and attempted curative surgery for locally advanced oesophageal and oesophagogastric junctional cancer at the Royal Infirmary of Edinburgh between 2010 and 2012. Local institutional ethical and research and development approvals were in place (REC references 06/S1101/16 and 10/S1402/33) (R&D ID 2006/W/PA/01). All patients gave informed consent and participants and their donated samples were de-identified at the time of recruitment. Patients were selected for relative clinical homogeneity with respect to known prognostic variables including lymphatic metastasis and tumour differentiation¹⁸. The clinical characteristics of the cohort are presented in Table 2.

At the commencement of this study, no shotgun proteomic data were available for oesophageal adenocarcinoma tissue to inform a power calculation for sample size determination. The sample number was therefore based on previous oesophageal discovery-phase proteomic studies or studies in similar tissue types¹⁹⁻²², and the availability of high quality clinical material.

Due to the risk of false-positives due to the small sample size, proposed tumour-specific proteins identified by mass-spectrometry were additionally verified by immunohistochemistry (IHC) using cores from archival tumours and matched normal and metastatic tissues from an independent cohort of 115 patients with oesophageal or OGJ adenocarcinoma (clinical characteristics in Supplementary Table 1).

Sample processing

The sample processing workflow is summarised in Figure 1. Fresh tissue biopsies were snap frozen within 30 minutes of tumour extirpation and maintained in liquid nitrogen or on dry ice until lysis.

Frozen sections from each biopsy were reviewed by a consultant histopathologist to confirm the histological diagnosis and, for tumour biopsies, a minimum of 50% tumour cellularity.

The published Filter-Aided Sample Preparation (FASP) method was adapted for protein extraction and tryptic digestion from oesophagogastric tissue ²³.

Biopsies between 30mg and 60mg in weight were maintained on dry ice until rapid disruption at room temperature (RT) in low-binding micro-centrifuge tubes containing 1mm ceramic beads (Matrix D, MP Bio) by rapid shaking in a bench-top homogenizer (FastPrep-24, MP Bio) for 40 seconds at 6ms⁻¹.

Homogenates were dissolved in FASP lysis buffer (4% w/v SDS, 100 mM Tris/HCl, 100 mM DTT, pH 7.6), mixed for 20 minutes at RT and sonicated at maximum amplitude, for 30 seconds, on ice, using a needle sonicator (Bioruptor). Sonicated lysates were heated for 5 minutes at 95°C and clarified by centrifugation at 14,000g for 5 minutes at 20°C before buffer-exchange as per the published FASP protocol ²³. Trypsinisation was performed off-column, overnight at 37°C at a ratio of 100:1 (lysate protein mass : trypsin mass) using sequencing grade modified trypsin (Promega) as per manufacturer's instructions. Protein concentration was determined using a modified Lowry procedure as per the manufacturer's recommendations (RC-DC, Biorad).

Isobaric labelling and Fractionation of Tryptic peptides

Tryptic peptides from each tissue sample were independently labelled with one of the 6 Tandem Mass Tag (TMT) reagents (ThermoScientific) in technical duplicate per manufacturer's instructions (Figure 1). Labelled peptides from a single patient (6 reporter ions) were pooled, desalted on a Macro SpinColumn C18 (Harvard Apparatus) and separated into 24 fraction by OFFGEL electrophoresis as previously described ²⁴.

Liquid Chromatography

Dried, desalted peptide fractions were reconstituted in 5% (v/v) acetonitrile, 0.1% (v/v) Formic Acid (FA) in dH₂O and ~0.5 µg loaded onto a homemade, 100 µm internal diameter, 20 mm long trapping column packed with 200 Å, 5 µm Magic C18 AQ (Michrom, Auburn, CA, USA). Trapped peptides were eluted into a 75 µm internal diameter, 150 mm long analytical column packed with 100 Å, 3 µm Magic C18 AQ (Microcom, Auburn, CA, USA).

For ultra-performance liquid chromatography (UPLC), peptides were separated using a variable solvent gradient created by a combination of 0.1% (v/v) FA in dH₂O (solvent A) and 0.1% (v/v) FA in acetonitrile (solvent B). The gradient was run as follows: 0–1 mins, 95% (A) and 5% (B), 1–56 mins, 65% (A) and 35% (B), 66–76 mins, 20% (A) and 80% (B) using a flow rate of 220 nL/min.

Mass Spectrometry

Peptides were analysed in positive ion mode after electrospray ionisation on an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific, San Jose, CA, USA). For MS survey scans, the Orbitrap (OT) resolution was set to 60,000 and the ion population was set to 5×10^5 with an m/z window from 400 to 2,000. A maximum of 3 precursor ions with the greatest peak intensities were selected for both collision-induced dissociation (CID) and high-energy C-trap dissociation (HCD) in the LTQ with analysis in the OT. For fragment ion analysis in the LTQ, the ion population was set to 7×10^3 (isolation width of 2 m/z) while for detection in the OT, the ion population was set to 2×10^5 (isolation width of 2.5 m/z), with resolution of 7,500, first mass at m/z = 100, and maximum injection time of 750 ms. The normalized collision energies were set to 35% for CID and 60% for HCD.

Protein Identification

Protein identifications were made using the Easyprot platform (v2.3 build 720, Swiss Institute of Bioinformatics)²⁵. Data manipulation was performed using Excel (Version 14.0.6129.5000, Microsoft Office Professional 2010), R (version 2.15.1, General Public License), and custom scripts written in Perl (version 5.18.0, General Public License).

Thermo RAW files were converted to peak lists using ReAdW (version 4.3.1, ThermoFinnigan) and CID and HCD spectra were merged for simultaneous identification and quantification as previously described²⁶. Peaklist files were searched against the Uniprot human reference proteome (release 09/01/2013, containing 87,613 entries) using Phenyx® (version 2.6.1, GeneBio)²⁷ with a precursor

ion tolerance of 10 parts per million and a fragment ion tolerance of 0.6 Da. Variable peptide modifications included TMT-modified N-termini and lysines (additional 229.1629 Da) and oxidized methionines, with carbamidomethylation of cysteines set as a fixed modification. Trypsin was selected as the digestion enzyme, with one potential missed cleavage and a minimum of a single-tryptic terminus, a peptide length of 6 amino acids and a z-score of 4 were required.

All datasets were searched separately, once using a forward and once using a reversed protein database. The peptide false discovery rate was set to 1%. A single unique peptide was accepted for protein identification. Identified peptide sequences, scores, precursor m/z, corresponding proteins, protein coverage and raw data files have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository ²⁸ with the dataset identifier PXD004962.

Protein Quantitation

For relative protein quantitation, the TMT reporter ion intensities were extracted for each peptide. An isotopic purity correction was performed within Easyprot for each reporter based on the isotopic distribution of the sixplex-TMT reporters provided by the manufacturer.

The ratios of peptide expression between OAC and normal oesophagus (TvO) and OAC and normal gastric epithelium (TvG) were calculated for each peptide as the ratio of 126/128 reporter ion intensities and 126/130 reporter intensities respectively. This was repeated for 127/129 (TvO) and 127/131 (TvG) reporter ions as a technical replicate. The geometric mean peptide TvO and TvG ratios were calculated for each protein to derive an estimate of the relative protein expression between patient-matched tissue types, and limiting the skew introduced by outlier ratios ²⁹. Due to the ambiguity from potentially shared peptides, all protein isoforms were grouped under the parent protein identifier (Uniprot accession). The variance of each ratio and the number of peptides contributing to the mean were used for subsequent significance calculations.

Deriving a Mean Expression Ratio across Replicates

A mean expression value was derived across pooled technical and biological replicates by a meta-analysis approach using a fixed-effect model³⁰. The inverse of the variance in peptide reporter ratios was used to weight the contribution of protein expression from a replicate to the mean. In this manner, replicates with many peptides detected with similar relative tissue expression (low variance) for a protein contribute a greater proportion to the mean expression value (see Supplementary Methods). The un-weighted arithmetic mean and variance of the ratios across experiments was calculated for proteins identified by a single unique peptide in multiple replicates. The log₂-transformed protein expression ratios (TvO and TvG) were median normalised to account for systematic errors during TMT labelling such as minor variations in protein loading.

Statistical Tests

The inter-tissue ratios (TvO and TvG) were compared to technical replicates (TvT, OvO and TvT, GvG) using Welch's modified t-test³¹ as the variances between them differed significantly³²; Fligner-Killeen test $p < 10^{-10}$, Supplementary Figure 1. Welch's t-test was used to test the hypothesis that the relative protein expression between different tissues was not different from relative protein expression between technical replicates with "n" defined as the number of peptides contributing to the mean ratio (see Supplementary Methods)³³. All p-values were corrected using the Benjamini-Yekutieli method to control for multiple hypothesis testing³⁴. Significance was defined as a false discovery rate (FDR)-corrected $p < 0.05$ and p-values were two-tailed. Relative quantitation and significance are provided for all identified proteins (uploaded supplementary file; All_quantitation.xlsx).

Immunohistochemistry

For verification of relative protein expression, 4 µm sections were cut from formalin-fixed, paraffin-embedded (FFPE) blocks derived from the same resection specimens used to collect the fresh tissue for this study and representing normal oesophagus, normal stomach and OAC. Sections were subjected to immunohistochemical staining using standard techniques as previously described³⁵. Staining conditions were optimised for each antibody, included a no primary antibody control for each protein and are detailed in Supplementary Table 2.

Tissue Microarray

After institutional approval (R&D ID 2006/W/PA/01), FFPE blocks comprising normal oesophagus, normal stomach, OAC, normal lymph nodes and, if present, lymph node metastasis were identified for an independent cohort of 115 patients undergoing oesophageal resection for OAC between 1994 and 2005 (Supplemental Table 1). Representative cores (0.6mm) were transferred to a separate paraffin block as a tissue microarray (TMA). TMA sections (4 μ m) were stained as previously and scored using a modified Allred method ³⁶ by an expert histopathologist. Intensity was graded 0-3 (0 = nil, 1 = weak, 2 = moderate, 3 = strong) and frequency was graded 0-4 (0 = 0%, 1 = 1-10%, 2 = 10-50%, 3 = 50-80%, 4 = 80-100%). The sum of intensity + frequency was calculated for each scorable core. Not all cores could be scored due to loss of material during IHC or lack of the appropriate tissue type in the core.

Western Blotting and RNA interference

Lysates from each of the oesophageal cell types growing under basal conditions were resolved by SDS-PAGE (20 μ g protein per lane, 12% SDS gel) as previously described ³⁷. Western blots were probed with primary antibodies directed against ARHGDIB (ab88317, Abcam, 1/500 overnight at 4°C) or β -Actin (AC-15, Sigma, 1/5000, 2 hours at RT) followed by secondary incubation with rabbit anti-mouse or goat anti-rabbit antibodies conjugated to either the IRDye™ 680RD or IRDye™ 800CW ((Li-Cor Biosciences, 1/2000, 2 hours at RT protected from light). Blots were imaged using the Odyssey SA system (Li-Cor Biosciences) as per manufacturer's recommendations.

For siRNA experiments, OE33 cells were transfected with vector, non-targeting scrambled sequence siRNA (siGENOME, Non-Targeting siRNA#3, Dharmacon) or siRNA to ARHGDIB (siGENOME, SMARTPool, Dharmacon) as previously described ³⁷. Cells were harvested 72 hours after transfection and lysates resolved by western blotting as previously.

RESULTS

A total of 6,349 proteins were identified and quantified across all samples corresponding to 4,772 unique Entrez GeneIDs with 744 proteins quantified in both replicates from all seven samples. The protein identifications per patient are shown in Table 3.

Reproducibility of Quantitation

The reproducibility of quantification of protein expression was assessed by comparing technical replicates. Expression levels were highly correlated between technical replicates from the same patient's tissues (Figure 2A, B, median Pearson correlation coefficient (PCC) = 0.9811, $p < 0.001$).

There was also a reassuringly good correlation between different patients (biological replicates) when TvO ratios were considered (Figure 2C, median PCC = 0.555, $p < 0.001$) demonstrating concordance of the protein expression from histologically similar tissues within a relatively clinically homogeneous patient cohort. As expected, there was no significant correlation between TvO and TvG ratios across patients (Figure 2C, median PCC = 0.0115, $p > 0.05$) underscoring the diversity in protein expression of the tissues studied.

A Map of Protein Expression across Oesophagogastric Tissues

Protein expression was quantified by the ratio of TMT reporter ion intensities. Ratios were not, however, calculable for proteins exclusively expressed in one tissue and these proteins may have been excluded from our analysis despite their biological importance. Several peptides were manually identified with no reporter ion expression from one or two tissue types. In the context of other unique peptides identified from the same protein, however, no proteins with entirely tissue-specific expression could be identified in this study.

For the 4,181 proteins identified by more than one peptide in more than one replicate, and the 2,154 proteins identified by more than one peptide in a single replicate or a single peptide in more than one replicate a mean expression value and variance were derived (uploaded supplementary file; All_quantitation.xlsx). Those 14 proteins identified by a single peptide and only in a single replicate were considered low confidence identifications and were excluded from further analysis. The expression ratios for the 3,082 proteins significantly dysregulated between tissues (FDR-corrected

$p < 0.05$ for either TvG or TvO ratios) were used to produce a two-dimensional protein expression map with vectors of TvG expression on the x-axis and TvO on the y-axis (Figure 3).

Proteins over-expressed in OAC would be expected to have both high TvO and TvG ratios and therefore be identified in the upper right quadrant of the protein expression plot, discrete from other non-specifically expressed proteins. It was proposed that other proteins with tissue-specific expression patterns would also be closely associated on the plot. To test this hypothesis, proteins with an established expression pattern were considered. Gastric Intrinsic Factor (GIF) and Mucin 5AC are known to be specifically highly expressed in gastric epithelium^{38,39}. Reassuringly, both proteins were closely associated on the protein expression map and demonstrated similar expression in OAC and normal oesophageal tissue but high expression in gastric epithelium (Figure 3). Similarly, Keratins 4, 5 and 14 are known to be highly expressed in oesophageal squamous epithelium^{40,41} and were all clustered together. It was possible that the tissue expression pattern of other proteins could be inferred from their location on this map.

To test this, a further 2 proteins predicted to be upregulated in squamous tissue (heat shock protein family B (small) member 1; HSPB1 and transglutaminase 3; TGM3), two proteins upregulated in tumour but also generally expressed (SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1; SAMHD1 and Rho GDP dissociation inhibitor beta; ARHGDIB), two proteins upregulated in both tumour and gastric tissue (anterior gradient 2, protein disulphide isomerase family member; AGR2 and heat shock protein family A (Hsp70) member 5; HSPA5) and one protein predicted to be specifically highly expressed in tumour (epithelial cell adhesion molecule; EPCAM) (Supplementary Table 3) were selected for verification by immunohistochemistry (IHC).

Verification of Protein-expression differences across OAC and patient-matched normal tissues.

Sections from the FFPE tissue blocks from the original resection specimens used to derive the fresh tissue samples for proteomic analysis were subjected to IHC (Figure 4). All five of the proteins predicted to be upregulated in OAC compared to normal oesophagus (SAMHD1, ARHGDIB, AGR2, HSPA5, EPCAM) showed higher expression in the tumour sections compared to squamous epithelium. Similarly, the proteins predicted to be expressed preferentially in squamous epithelium;

HSPB1 and TGM3, showed the highest expression in normal squamous oesophageal tissue (Supplementary Table 3).

Although false positive homogeneous cytoplasmic staining was observed in all or scattered basal crypt epithelial cells in most gastric tissue sections (Figure 4; No primary antibody)⁴², both HSPA5 and AGR2 demonstrated true gastric epithelial staining as expected and were overexpressed in OAC compared to squamous epithelium.

SAMHD1 showed more widespread nuclear expression across tissues but was mildly upregulated in OAC cells, as predicted. In contrast, EPCAM was very highly expressed in tumour cells with only moderate staining in basal epithelial squamous cells and gastric epithelial cells. Although overall expression of ARHGDIB was indeed higher in OAC compared to normal squamous and gastric tissue, this staining was observed in stromal cells, most likely lymphocytes, rather than epithelial-derived tumour cells.

These findings support the accuracy of the quantitative proteomic approach for each of the seven candidates selected. As expected from the proteomic data, EPCAM demonstrated the greatest specificity for tumour cells.

To determine the specificity of EPCAM for OAC cells compared to surrounding normal tissues, protein expression was determined by IHC using a tissue microarray consisting; normal gastric tissue, normal squamous tissue, uninvolved lymph nodes, involved lymph nodes and primary tumour samples from resection specimens from 115 patients whom had undergone surgical resection for OAC (Figure 5). EPCAM was expressed at low levels in basal squamous epithelial cells and low to moderate levels in gastric epithelium. In contrast EPCAM was highly expressed in OAC and was expressed at higher levels than the median normal gastric or normal oesophageal epithelial expression in 98% of tumours. This high specificity was demonstrated in metastatic lesions as well as primary tumours with no EPCAM expression detectable in normal lymph nodes but high expression in 93% of lymph node metastases.

This proteomic strategy has therefore demonstrated the ability to detect relatively specific markers of OAC. Those proteins predicted to be highly expressed in OAC compared to surrounding normal tissues, including several novel candidate therapeutic targets, are presented in more detail in Figure 6.

DISCUSSION

There is a need to identify selective markers of oesophageal adenocarcinoma for early diagnosis, to enhance clinical staging and direct novel therapies. This shotgun proteomic study compared protein expression from matched oesophageal adenocarcinoma, normal oesophagus and normal gastric samples from 7 patients and provides quantitative data on protein expression for over 6,000 proteins across these tissues. A comparative analysis approach was employed to select tumour-specific proteins and this method was verified to be accurate by IHC. Multiple novel tumour-specific proteins are proposed and EPCAM was demonstrated to be specifically over-expressed in primary tumours and lymph node metastases.

Incomplete proteome coverage remains a significant limitation of all proteomic studies, however, and only 744 proteins (12%) were detected in all technical replicates across the seven patients' tissues. Despite this intrinsic limitation of proteomic studies, the technical reproducibility of protein quantitation was very high and a strategy was developed to identify significantly dysregulated proteins.

The combination of TvO and TvG ratios for each quantified protein allowed the generation of a 2D expression map. The relative quantitative accuracy of this approach was confirmed first by evaluation of proteins with a well validated expression profile (GIF, Muc5AC, Keratin 4,5,14) and then subsequently by IHC for 7 further proteins with varied expression profiles (ARHGDIB, SAMHD1, AGR2, HSPA5, EPCAM, TGM3, HSPB1). In each case the observed staining pattern mirrored the expected expression from the proteomic data (Supplementary Table 3). This provides confidence in the predicted expression for proteins with an unknown pattern.

Both TGM3 and HSPB1 were found to be expressed at high levels in squamous epithelium compared to OAC, as expected from our proteomic data and as reported in previous work ⁴³. In contrast, AGR2 was found to be expressed in gastric epithelium and OAC with no expression observed in squamous epithelia. Similarly, HSPA5 was found to be expressed in both OAC and gastric epithelium. These findings agree with previous reports ⁴⁴⁻⁴⁸. HSPA5 is thought to play a key role in the regulation of the unfolded-protein response and this expression pattern may reflect an increased protein chaperone demand in secretory cells ^{48,49}. Indeed AGR2 has a function in protein homeostasis and secretion ⁵⁰.

SAMHD1 exhibited nuclear expression in both epithelial and non-epithelial cells with the highest expression in OAC. It has a reported role in restricting HIV replication and modulating the immune response in T cells ⁵¹ but there is limited data for its role in cancer and this would be worthy of further study.

ARHGDIB was over-expressed in OAC sections compared to normal squamous and normal gastric tissue. The ARHGDIB positive cells appeared, however, to be lymphocytes rather than epithelial-derived tumour cells. This expression pattern has been observed for ARHGDIB with a different antibody ⁴⁷ and a multi-tissue study suggested expression was restricted to haematopoietic cells ⁵².

Divergent roles for ARHGDIB have subsequently been proposed in the literature with some evidence for a role in the suppression of metastasis in bladder cancer ⁵³ and in contrast a pro-invasive role in gastric cancer ⁵⁴. One previous proteomic study reported ARHGDIB overexpression in OAC compared to Barrett's epithelium at both the mRNA and protein level²². In contrast to our study, the previous work presented cytoplasmic and membrane staining in epithelial cells with minimal stromal positivity in OAC sections. The specificity of the rabbit anti-ARHGDIB antibody used in that study was not demonstrated in the manuscript. In contrast the antibody in this study identifies a protein of the predicted mass (48kDA) in a panel of oesophageal cell lines by western blotting and is specific for that protein as confirmed by siRNA (Supplementary Figure 2). The variable expression of ARHGDIB noted in the panel of oesophageal cell lines may reflect context-dependent regulation. Staining across a larger number of oesophageal tumours could establish greater confidence over the cell-specific expression profile.

The cancer antigen EPCAM was predicted to be highly expressed in OAC cells compared to surrounding normal tissues and this was indeed observed. This has previously been demonstrated for several cancer types, including oesophageal ^{55,56}. EPCAM is a cell adhesion molecule that is highly expressed on the cell membrane and may possess a signalling role through the regulation of cell proliferation via a cleaved intracellular domain ⁵⁷. Due to its high specificity for malignancy, EPCAM-based assays have been developed to detect circulating tumour cells although these are less sensitive for mesenchymal tumours ⁵⁸.

The expression of EPCAM was examined in detail by IHC in an independent cohort of 115 oesophageal adenocarcinomas with matched normal and metastatic tissues. EPCAM was highly

expressed in the clear majority of OACs with higher EPCAM histoscores observed in 85/87 OACs than the median gastric or oesophageal scores. EPCAM was also highly specifically expressed in lymph node metastases compared to surrounding normal lymph nodes raising the possibility that it could be exploited to enhance clinical staging using novel techniques ⁵⁹.

These data agree with previous work demonstrating over-expression of EPCAM in OAC compared to surrounding normal tissues ⁵⁶. A further study identified disseminated tumour cells from bone marrow and lymph nodes in patients with oesophageal cancer and, although the primary tumours predominantly expressed high levels of EPCAM, EPCAM expression was only observed in 37% of disseminated tumour cells from bone marrow aspirates ⁶⁰. A reduction in EPCAM expression by RNA interference increased migration in vitro and the authors proposed that EPCAM expression reduced during the process of invasion as cells adopted a more mesenchymal phenotype. This may have clinical implications if anti-EPCAM therapies are to be considered.

Samples were obtained from six men and one woman (Table 2) reflecting the four-fold greater prevalence of OAC in men ¹⁸. Although this bias could limit the applicability of this study to women with OAC, no differences were noted in the expression patterns of EPCAM across tissue types from men (n=97) and women (n=18) in the TMA.

OAC exhibits a high frequency of DNA mutation and mutant proteins are highly likely to be tumour specific ¹². A limitation of this study is that mass spectra were searched against a protein database containing only wildtype proteins, however, so that mutant proteins could not be identified. An alternative strategy is to generate mutant protein databases using tumour genome sequencing data, either specific to the patient's tumour or using commonly identified variants and search mass spectra against these ⁶¹. This proteogenomic approach has the potential to reveal tumour-specific proteins, however, significant technical challenges remain in controlling the protein database size and false-discovery rate ⁶².

Potentially due to this limitation, no entirely tumour-specific proteins could be identified in this study. Importantly, however, a group of proteins highly expressed in tumours relative to surrounding normal tissues was proposed (Figure 6). Immunotherapeutic trials are already underway in other cancer types with agents directed against several of these including EPCAM ⁶³, glycoprotein A33; GPA33 ^{64,65}, mucin 1, cell surface associated; MUC1 ⁶⁶ and melanoma antigen; MAGE ⁶⁷ proteins. If the

expression of these can be validated to be specific to OAC cells over surrounding tissues, there would be a compelling rationale to expand trials to include patients with OAC and to develop specific imaging tools to these targets to enhance clinical staging.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. (2011) Global cancer statistics. *CA Cancer J Clin.* 61(2):69-90.
2. Lepage C, Rachet B, Jooste V, Faivre J, Coleman MP. (2008) Continuing rapid increase in esophageal adenocarcinoma in England and Wales. *Am J Gastroenterol.* 103(11):2694-2699.
3. Pohl H, Welch H. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. (2005) *J Natl Cancer Inst.* 97(2):142-146.
4. Groene O, Cromwell D, Hardwick R, Riley S, Crosby T, Greenaway K. (2012) *National Oesophago-Gastric Cancer Audit*. Royal College of Surgeons of England.
5. Allum WH, Blazeby JM, Griffin SM, et al. (2011) Guidelines for the management of oesophageal and gastric cancer. *Gut.* 60(11):1449-1472.
6. O'Neill JR, Kennedy ED, Save V, et al. (2017) Patients unfit for neoadjuvant therapy may still undergo resection of locally advanced esophageal or esophagogastric junctional cancer with acceptable oncological results. *IJS Oncology.* 2(2):e09.
7. Dikken JL, van Sandick JW, Allum WH, et al. (2013) Differences in outcomes of oesophageal and gastric cancer surgery across Europe. *Br J Surg.* 100(1):83-94.
8. Fischer C, Lingsma H, Hardwick R, Cromwell DA, Steyerberg E, Groene O. Risk adjustment models for short-term outcomes after surgical resection for oesophagogastric cancer. (2016) *Br J Surg.* 103(1):105-116.
9. McElroy M, Kaushal S, Luiken GA, et al. Imaging of primary and metastatic pancreatic cancer using a fluorophore-conjugated anti-CA19-9 antibody for surgical navigation. (2008) *World J Surg.* 32(6):1057-1066.
10. Souza RF, Krishnan K, Spechler SJ. Acid, bile, and CDX: the ABCs of making Barrett's metaplasia. (2008) *Am J Physiol Gastrointest Liver Physiol.* 295:G211-218.

11. Gertler R, Stein HJ, Langer R, et al. Long-term outcome of 2920 patients with cancers of the esophagus and esophagogastric junction: evaluation of the New Union Internationale Contre le Cancer/American Joint Cancer Committee staging system. (2011) *Ann Surg.* 253(4):689-698.
12. Weaver JM, Ross-Innes CS, Shannon N, et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. (2014) *Nat Genet.* 46(8):837-843.
13. McGranahan N, Furness AJ, Rosenthal R, et al. (2016) Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science.* 351(6280):1463-1469.
14. Rosenberg SA, Yang JC, Sherry RM, et al. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. (2011) *Clin Cancer Res.* 17(13):4550-4557.
15. Singhal R, Carrigan JB, Wei W, et al. MALDI profiles of proteins and lipids for the rapid characterisation of upper GI-tract cancers. (2013) *J Proteomics.* 80C:207-215.
16. Geiger T, Velic A, Macek B, et al. Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. (2013) *Mol Cell Proteomics.* 12(6):1709-1722.
17. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. (2014) *Nature.* 509(7502):575-581.
18. O'Neill JR, Stephens NA, Save V, et al. Defining a positive circumferential resection margin in oesophageal cancer and its implications for adjuvant treatment. (2013) *Br J Surg.* 100(8):1055-1063.
19. Cai Z, Zhao JS, Li JJ, et al. A combined proteomics and metabolomics profiling of gastric cardia cancer reveals characteristic dysregulations in glucose metabolism. (2010) *Mol Cell Proteomics.* 9(12):2617-2628.
20. Besson D, Pavageau AH, Valo I, et al. A quantitative proteomic approach of the different stages of colorectal cancer establishes OLFM4 as a new nonmetastatic tumor marker. (2011) *Mol Cell Proteomics.* 10(12):M111.009712.
21. Peng D, Sheta EA, Powell SM, et al. Alterations in Barrett's-related adenocarcinomas: A proteomic approach. (2008) *International Journal of Cancer.* 122:1303-1310.

22. Zhao J, Chang AC, Li C, et al. Comparative proteomics analysis of Barrett metaplasia and esophageal adenocarcinoma using two-dimensional liquid mass mapping. (2007) *Mol Cell Proteomics*. 6(6):987-999.
23. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. (2009) *Nat Methods*. 6(5):359-362.
24. Dayon L, Turck N, Kienle S, et al. Isobaric tagging-based selection and quantitation of cerebrospinal fluid tryptic peptides with reporter calibration curves. (2010) *Anal Chem*. 82(3):848-858.
25. Gluck F, Hoogland C, Antinori P, et al. EasyProt--an easy-to-use graphical platform for proteomics data analysis. (2013) *J Proteomics*. 79:146-160.
26. Dayon L, Pasquarello C, Hoogland C, Sanchez JC, Scherl A. Combining low- and high-energy tandem mass spectra for optimized peptide quantification with isobaric tags. (2010) *J Proteomics*. 73(4):769-777.
27. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J. OLAV: towards high-throughput tandem mass spectrometry data identification. (2003) *Proteomics*. 3(8):1454-1463.
28. Vizcaíno JA, Côté RG, Csordas A, et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. (2013) *Nucleic Acids Res*. 41(Database issue):D1063-1069.
29. Shaffer J. Caution on the Use of Variance Ratio: A Comment. (1992) *Review of Educational Research*. 62(4):429-432.
30. Borenstein M. *Introduction to meta-analysis*. (2009) Chichester, U.K.: John Wiley & Sons.
31. Ruxton G. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. (2006) *Behavioural Ecology*. 17(4):688-690.
32. Conover W, Johnson M, Johnson M. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. (1981) *Technometrics*. 23(4):351-361.
33. Welch B. The generalization of "Student's" problem when several different population variances are involved. (1947) *Biometrika*. 34(1-2):28-35.
34. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. (2001) *Annals of Statistics*. 29(4):1165-1185.

35. Murray E, Hernychová L, Scigelova M, et al. Quantitative proteomic profiling of pleomorphic human sarcoma identifies CLIC1 as a dominant pro-oncogenic receptor expressed in diverse sarcoma types. (2014) *J Proteome Res.* 13(5):2543-2559.
36. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. (1998) *Mod Pathol.* 11(2):155-168.
37. Huart AS, MacLaine NJ, Meek DW, Hupp TR. CK1alpha plays a central role in mediating MDM2 control of p53 and E2F-1 protein stability. (2009) *J Biol Chem.* 284(47):32384-32394.
38. Freeman TC, Ivens A, Baillie JK, et al. A gene expression atlas of the domestic pig. (2012) *BMC Biol.* 10:90.
39. Goebel M, Stengel A, Lambrecht NW, Sachs G. Selective gene expression by rat gastric corpus epithelium. (2011) *Physiol Genomics.* 43(5):237-254.
40. Takashima M, Kawachi H, Yamaguchi T, et al. Reduced expression of cytokeratin 4 and 13 is a valuable marker for histologic grading of esophageal squamous intraepithelial neoplasia. (2012) *J Med Dent Sci.* 59(1):17-28.
41. Wang X, Ouyang H, Yamamoto Y, et al. Residual embryonic cells as precursors of a Barrett's-like metaplasia. (2011) *Cell.* 145(7):1023-1035.
42. Ramos-Vara JA. Principles and methods of immunohistochemistry. (2011) *Methods Mol Biol.* 691:83-96.
43. Soldes OS, Kuick RD, Thompson IA, et al. Differential expression of Hsp27 in normal oesophagus, Barrett's metaplasia and oesophageal adenocarcinomas. (1999) *Br J Cancer.* 79(3-4):595-603.
44. Pohler E, Craig A, Cotton J, et al. The Barrett's antigen anterior gradient-2 silences the p53 transcriptional response to DNA damage. (2004) *Mol Cell Proteomics.* 3(6):534-547.
45. Pizzi M, Fassan M, Realdon S, et al. Anterior gradient 2 profiling in Barrett columnar epithelia and adenocarcinoma. (2012) *Hum Pathol.* 43(11):1839-1844.
46. Thompson DA, Weigel RJ. hAG-2, the human homologue of the *Xenopus laevis* cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines. (1998) *Biochem Biophys Res Commun.* 251(1):111-116.
47. Uhlén M, Björling E, Agaton C, et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. (2005) *Mol Cell Proteomics.* 4(12):1920-1932.

48. Langer R, Feith M, Siewert JR, Wester HJ, Hoeffler H. Expression and clinical significance of glucose regulated proteins GRP78 (BiP) and GRP94 (GP96) in human adenocarcinomas of the esophagus. (2008) *BMC Cancer*.8:70.
49. Hetz C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. (2012) *Nat Rev Mol Cell Biol*. 13(2):89-102.
50. Chevet E, Fessart D, Delom F, et al. Emerging roles for the pro-oncogenic anterior gradient-2 in cancer development. (2013) *Oncogene*.32(20):2499-2509.
51. Baldauf HM, Pan X, Erikson E, et al. SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. (2012) *Nat Med*.18(11):1682-1687.
52. Scherle P, Behrens T, Staudt LM. Ly-GDI, a GDP-dissociation inhibitor of the RhoA GTP-binding protein, is expressed preferentially in lymphocytes. (1993) *Proc Natl Acad Sci U S A*. 90(16):7568-7572.
53. Moissoglu K, McRoberts KS, Meier JA, Theodorescu D, Schwartz MA. Rho GDP dissociation inhibitor 2 suppresses metastasis via unconventional regulation of RhoGTPases. (2009) *Cancer Res*. 69(7):2838-2844.
54. Cho HJ, Baek KE, Park SM, et al. RhoGDI2 expression is associated with tumor growth and malignant progression of gastric cancer. (2009) *Clin Cancer Res*. 15(8):2612-2619.
55. Went PT, Lugli A, Meier S, et al. Frequent EpCam protein expression in human carcinomas. (2004) *Hum Pathol*. 35(1):122-128.
56. Kimura H, Kato H, Faried A, et al. Prognostic significance of EpCAM expression in human esophageal cancer. (2007) *Int J Oncol*.30(1):171-179.
57. Maetzel D, Denzel S, Mack B, et al. Nuclear signalling by tumour-associated antigen EpCAM. (2009) *Nat Cell Biol*. 11(2):162-171.
58. Rao CG, Chianese D, Doyle GV, et al. Expression of epithelial cell adhesion molecule in carcinoma cells present in blood and primary and metastatic tumors. (2005) *Int J Oncol*. 27(1):49-57.
59. Thorek DL, Ulmert D, Diop NF, et al. Non-invasive mapping of deep-tissue lymph nodes in live animals using a multimodal PET/MRI nanoparticle. (2014) *Nat Commun*. 5:3097.
60. Driemel C, Kremling H, Schumacher S, et al. Context-dependent adaption of EpCAM expression in early systemic esophageal cancer. (2014) *Oncogene*. 33(41):4904-4915.

61. Zhang B, Wang J, Wang X, et al. Proteogenomic characterization of human colon and rectal cancer. (2014) *Nature*. 513(7518):382-387.
62. Blakeley P, Overton IM, Hubbard SJ. Addressing statistical biases in nucleotide-derived protein databases for proteogenomic search strategies. (2012) *J Proteome Res*. 11(11):5221-5234.
63. Seimetz D, Lindhofer H, Bokemeyer C. Development and approval of the trifunctional antibody catumaxomab (anti-EpCAM x anti-CD3) as a targeted cancer immunotherapy. (2010) *Cancer Treat Rev*. 36(6):458-467.
64. Sakamoto J, Oriuchi N, Mochiki E, et al. A phase I radioimmunolocalization trial of humanized monoclonal antibody huA33 in patients with gastric carcinoma. (2006) *Cancer Sci*. 97(11):1248-1254.
65. Chong G, Lee FT, Hopkins W, et al. Phase I trial of 131I-huA33 in patients with advanced colorectal carcinoma. (2005) *Clin Cancer Res*. 11(13):4818-4826.
66. Quoix E, Lena H, Losonczy G, et al. TG4010 immunotherapy and first-line chemotherapy for advanced non-small-cell lung cancer (TIME): results from the phase 2b part of a randomised, double-blind, placebo-controlled, phase 2b/3 trial. (2016) *Lancet Oncol*. 17(2):212-223.
67. Vansteenkiste J, Zielinski M, Linder A, et al. Adjuvant MAGE-A3 Immunotherapy in Resected Non-Small-Cell Lung Cancer: Phase II Randomized Study Results. (2013) *J Clin Oncol*. 31(19):2396-2403.
68. Yoo C, Zhao J, Pal M, et al. Automated integration of monolith-based protein separation with on-plate digestion for mass spectrometric analysis of esophageal adenocarcinoma human epithelial samples. (2006) *Electrophoresis*. 27(18):3643-3651.
69. Langer R, Ott K, Specht K, et al. Protein expression profiling in esophageal adenocarcinoma patients indicates association of heat-shock protein 27 expression and chemotherapy response. (2008) *Clin Cancer Res*. 14(24):8279-8287.
70. Quaas A, Bahar AS, von Loga K, et al. MALDI imaging on large-scale tissue microarrays identifies molecular features associated with tumour phenotype in oesophageal cancer. (2013) *Histopathology*. 63(4):455-62.

71. Aichler M, Elsner M, Ludyga N, et al. Clinical response to chemotherapy in oesophageal adenocarcinoma patients is linked to defects in mitochondria. (2013) *J Pathol.* 230(4):410-419.
72. Elsner M, Rauser S, Maier S, et al. MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma. (2012) *J Proteomics.* 75(15):4693-4704.
73. Streitz JM, Madden MT, Marimanikkuppam SS, Krick TP, Salo WL, Aufderheide AC. Analysis of protein expression patterns in Barrett's esophagus using MALDI mass spectrometry, in search of malignancy biomarkers. (2005) *Dis Esophagus.* 18(3):170-176.

FOOTNOTES

Contributions – The individual author contributions (initials) are as listed.

Concept and design - JRON, AS, BV, IO, TRH. Tissue and clinical data collection and pathology review – JRON, VS, SPB, RN. Tissue processing, Mass Spectrometry and Immunohistochemistry – JRON, HSP, VS, RN, BV, AS. Data Analysis – JRON, EPC, SPB, BV, IO, TRH. Manuscript draft and revision – JRON, HSP, IO, TRH. Final approval - JRON, HSP, EPC, VS, SPB, RN, BV, IO, AS, TRH

The authors wish to acknowledge the support of the Edinburgh Experimental Cancer Medicine Centre for assistance with tissue collection, storage and tissue microarray construction and Tammy Piper and Professor John Bartlett for assistance with immunohistochemistry.

Conflict of Interest - The authors declare no conflict of interest.

These data have been presented, in part, at the Human Proteome Organisation (HUPO) World Congress in 2012 and the Association of Upper-GI Surgeons of Great Britain and Ireland (AUGIS) in 2013 and published in abstract form (British Journal of Surgery 2013;100(S8):55).

Funding - This work was funded in part by a Wellcome Trust Clinical PhD Fellowship, grant number 094417/Z/10/Z and by the project MEYS – NPS I – LO1413. BV and TH were supported with grant from the Czech Science Foundation GA CR 16-07321S. IO was supported by a Royal Society of Edinburgh Scottish Government Fellowship co-funded by Marie Curie Actions and the UK Medical Research Council MC_UU_12018/25.

FIGURE LEGENDS

Figure 1. Summary experimental workflow. Tryptic peptides from histologically-verified, fresh frozen tissue biopsies were labelled with TMT reporters prior to OFFGEL electrophoresis and tandem mass spectrometry. Abbreviations: ACC – Adenocarcinoma biopsies, Oes – normal squamous oesophageal biopsies, Sto – normal gastric biopsies. TMT – Tandem Mass Tags. LC – Liquid Chromatography, MS/MS – Tandem Mass Spectrometry.

Figure 2. Correlation of technical replicates. (A) Median-normalised \log_2 ratios are displayed for two technical replicates of samples from patient 44. TvO = \log_2 ratio of expression between OAC and normal oesophagus. TvG = \log_2 ratio of expression between OAC and normal gastric tissue. (B) Box and whisker plot summarising the median (line within the box), 25th and 75th percentiles (box limits) and 5th and 95th percentiles (whiskers) for the Pearson's correlation coefficients (PCC) for technical replicates across all experiments (n=14). (C) Heatmap representation of PCC between technical replicates across all experiments. Unsupervised hierarchical clustering was performed using an agglomerative complete linkage method to generate the similarity dendrogram (left).

Figure 3. Distribution of Significantly Dysregulated Proteins. Significance was defined as an FDR-corrected $p < 0.05$ for the comparison of TvO and the mean of TvT and OvO ratios or for the comparison of TvG and the mean of TvT and GvG ratios. Each point on this figure represents a single protein that is significantly dysregulated in at least one tissue type. Proteins have been plotted according to their \log_2 expression ratios and points sized per the frequency of observation by MS across pooled biological and technical replicates (range 1-14). Selected proteins, labelled by their official gene names, are highlighted in green for proteins known to be specifically expressed in gastric tissue, purple for proteins known to be expressed in squamous tissue and red for proteins selected for further validation by IHC.

Figure 4. IHC staining pattern for selected proteins. Representative images are provided highlighting the staining pattern from OAC and matched normal tissues from the original resection specimens used to derive the fresh frozen biopsies.

Figure 5. EPCAM expression in oesophageal adenocarcinoma, lymph node metastases and corresponding normal tissues. Raw histoscores are presented as points with medians as dashed blue lines. Error bars in blue represent interquartile ranges. The number of scorable cores examined is presented below each column as well as a representative core image.

Figure 6. Annotation of Proteins Highly Expressed in Tumours Compared to Normal Oesophageal and Gastric Epithelium. Points are identical to Figure 3 but only proteins within a selected expression range are shown. Proteins have been annotated with their official gene names and are sized according to the frequency of observation by MS across pooled technical and biological replicates (range 1-14).

TABLES

Table 1. Published Proteomic Studies of OAC

Study Author	Patients	Method	Tissue Preparation	Squamous Samples	Gastric Samples	Barrett's Samples	ACC Samples	SCC Samples	Total Identifications	Number of Proteins Dysregulated
Zhao ²²	6	LC-ESI TOF MS, Targeted LC-MS/MS	Fresh frozen biopsies	-	-	6	6	-	-	38 proteins
Yoo ⁶⁸	1	LC-ESI TOF MS, Targeted MALDI-MS	Fresh frozen biopsies	-	-	-	1	-	22 proteins	-
Peng ²¹	8	2D Gels, Targeted MALDI-MS	Fresh frozen biopsies	2	2	-	8	-	-	23 dysregulated gel spots – 22 proteins identified.
Langer ⁶⁹	20	2D Gels, Targeted MALDI- and LC- MS/MS	Fresh frozen biopsies				20		-	Data for 4 proteins presented
Quaas ⁷⁰	477	MALDI-MSI	FFPE slides from a tissue microarray	-	-	-	300	177	72 spectral features 13 peptides	-
Aichler ⁷¹	23	MALDI-MSI. Targeted LC-MS/MS	Fresh frozen biopsies	-	-	-	23		-	22 spectral features, 6 proteins.

Identifying oesophageal adenocarcinoma-specific proteins

Elsner ⁷²	38	MALDI-MSI. Targeted LC-MS/MS	Fresh frozen biopsies	-	-	11	33			61 spectral features, 6 proteins
Streitz ⁷³	4	MALDI-MS	LCM of fresh frozen biopsies	-	-	4	4		8 spectral features	
Singhal ¹⁵	53 (iTRAQ)	MALDI-MS iTRAQ LC-MS/MS	Fresh frozen biopsies	30	23		30	-	972 proteins	Not described.

Abbreviations: OAC – Oesophageal Adenocarcinoma, ACC – Adenocarcinoma, SCC – Squamous Cell Carcinoma.

Table 2. Clinical Characteristics of Patients Donating Tissue for Proteomics Analysis

Patient (Pt)							
	Pt44	Pt46	Pt48	Pt51	Pt53	Pt60	Pt61
Gender	Male	Male	Male	Male	Female	Male	Male
Age	59	67	65	41	52	60	58
Histology	ACC	ACC	ACC	ACC	ACC	ACC	ACC
Location	OGJ Type II	Oes Lower	OGJ Type II	OGJ Type I	Oes Lower	Oes Lower	Oes Lower
Neoadjuvant Therapy	2xCF	2xCF	2xCF	2xCF	2xCF	2xCF	2xCF
Surgery	ILO	ILO	ILO	ILO	ILO	ILO	ILO
Tumour Diameter	38 mm	70 mm	40 mm	50 mm	83 mm	52 mm	35 mm
PRM and DRM	>1 mm	>1 mm	>1 mm	>1 mm	>1 mm	>1 mm	>1 mm
Distance to CRM	4.2 mm	0.0 mm	0.0 mm	0.3 mm	3.0 mm	1.0 mm	1.0 mm
Resection	R0	R1	R1	R1	R0	R0	R0
Differentiation	Moderate	Poor	Poor	Poor	Poor	Poor	Poor
LVI	Y	Y	Y	Y	N	Y	Y
Venous Invasion	N	N	N	Y	N	N	N
PNI	N	Y	Y	Y	N	Y	Y
T stage	ypT2	ypT4a	ypT3	ypT3	ypT2	ypT3	ypT3
N Stage	ypN1	ypN3	ypN3	ypN3	ypN1	ypN3	ypN2
Positive Nodes	2	8	16	7	1	7	3
Nodes Resected	27	18	28	28	23	21	37
AJCC Stage	IIB	IIIC	IIIC	IIIC	IIB	IIIC	IIIB
Mandard TRG	V	V	IV	V	IV	V	V
Alive at Analysis	No	No	No	No	Yes	No	Yes
Overall Survival	48.1 months	15.3 months	10.9 months	24.1 months	47.6 months (censored)	17.8 months	49.0 months (censored)
Recurrence-Free Survival	34.8 months	12.6 months	10.1 months	10.9 months	47.6 months (censored)	8.1 months	49.0 months (censored)

Abbreviations: ACC – Adenocarcinoma, 2xCF - 2 cycles of Cisplatin and 5-Fluorouracil, ILO – Ivor-Lewis Oesophagectomy, mm – millimetre, AJCC – American Joint Committee on Cancer, OGJ – Oesophagogastric Junctional Tumour, PRM – Proximal resection margin, DRM – Distal resection margin, CRM – Circumferential resection margin., LVI – Lymphovascular invasion, PNI – perineural invasion Y-Yes, N-No, CT – Computed Tomography, Mandard TRG – Tumour Regression Grade.

Table 3. Protein Identifications by Patient

Patient	Total Proteins (1% peptide FDR, 1 unique peptide per protein)	Unique to patient
Pt44	2,901	368
Pt46	2,534	256
Pt48	3,309	550
Pt51	3,327	503
Pt53	2,904	280
Pt60	2,369	220
Pt61	2,828	264

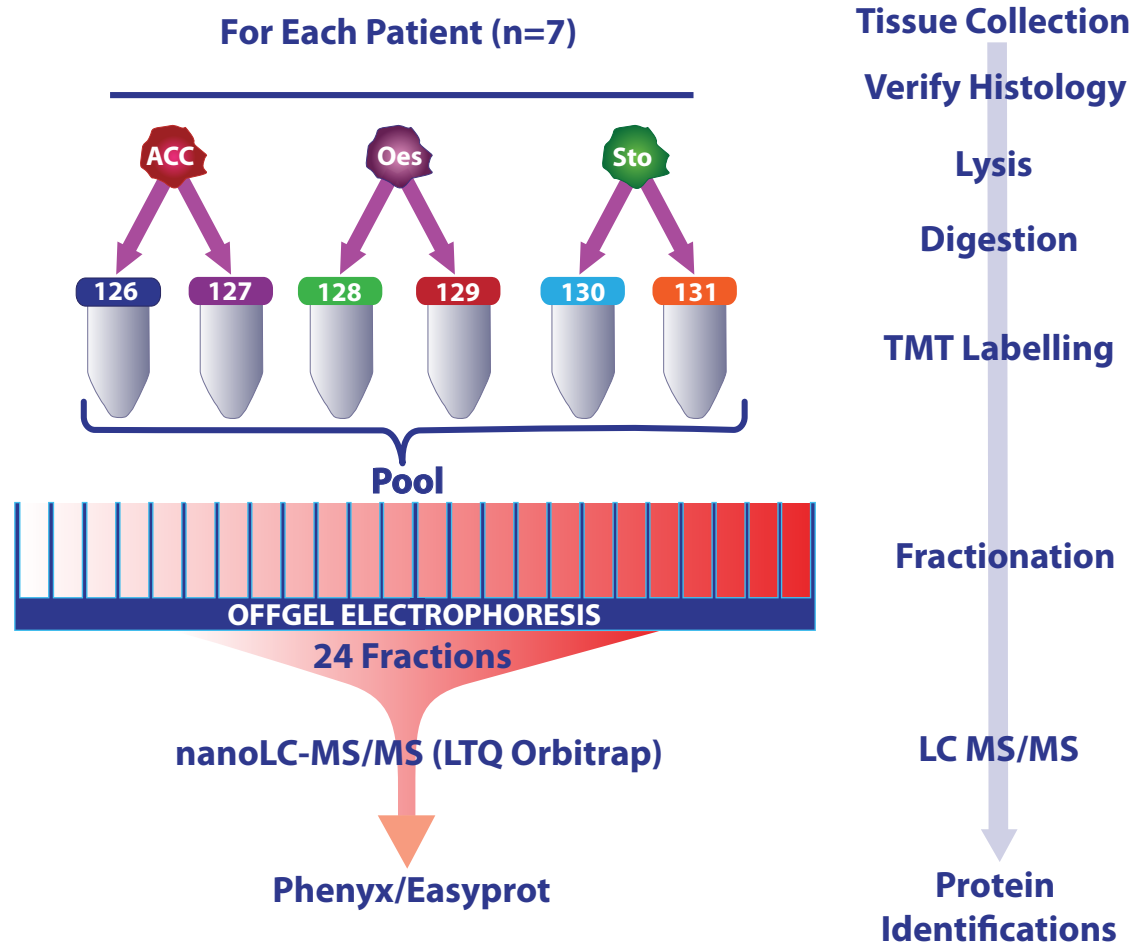
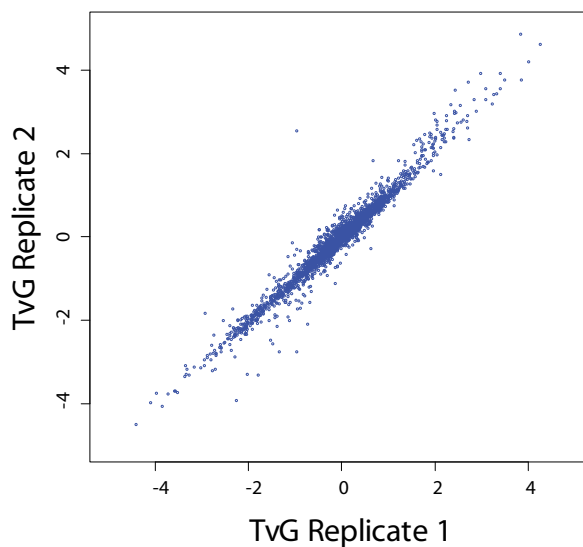
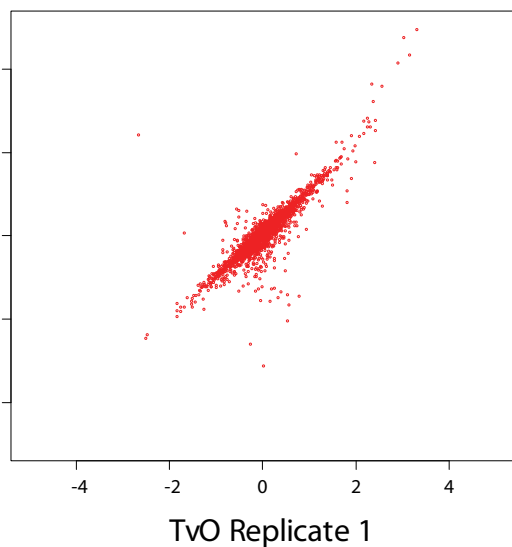


Figure 1

Patient 44 Median-Normalised Log₂ Ratios

B

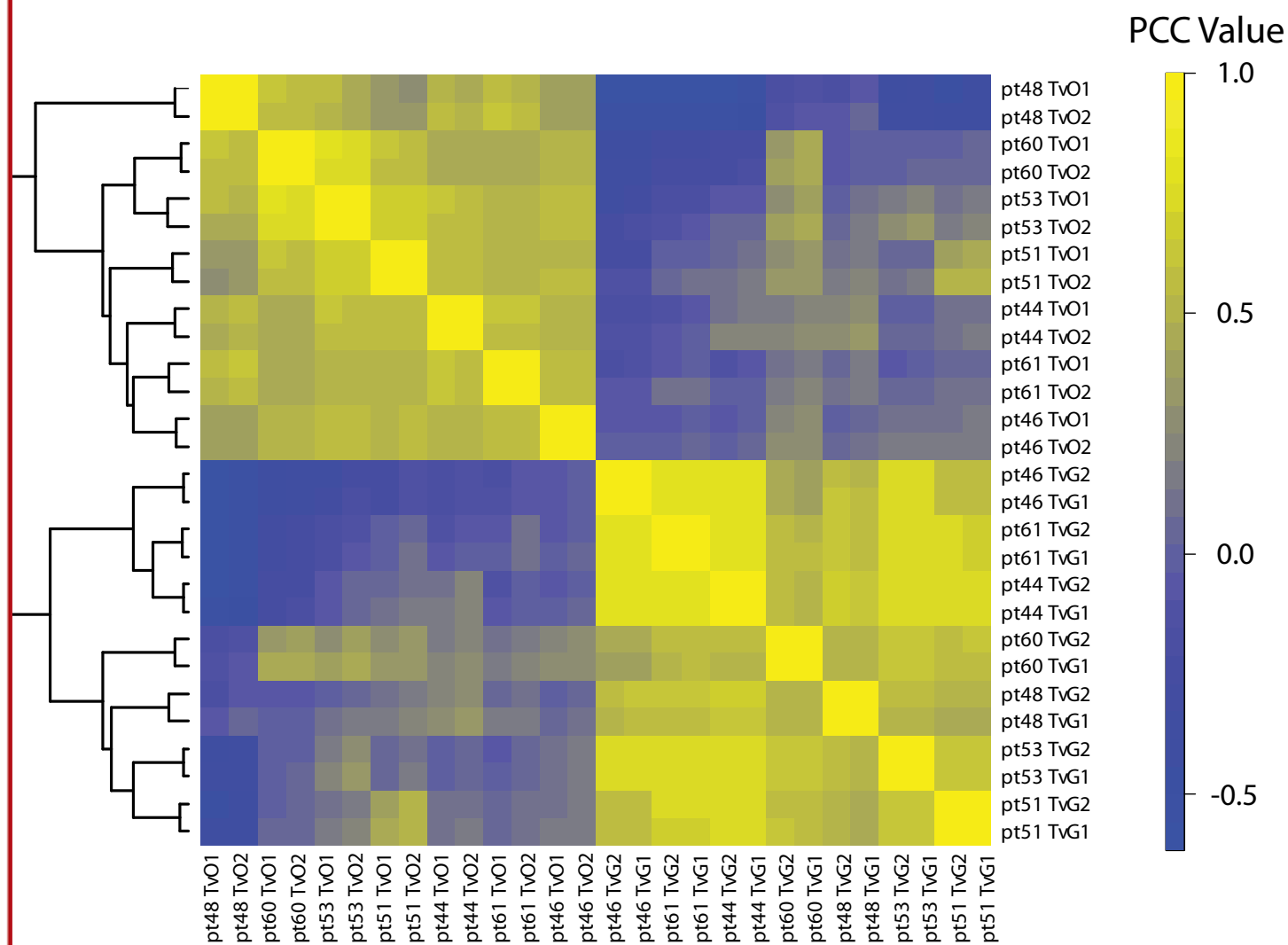
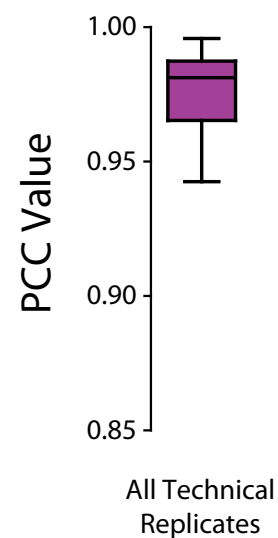


Figure 2

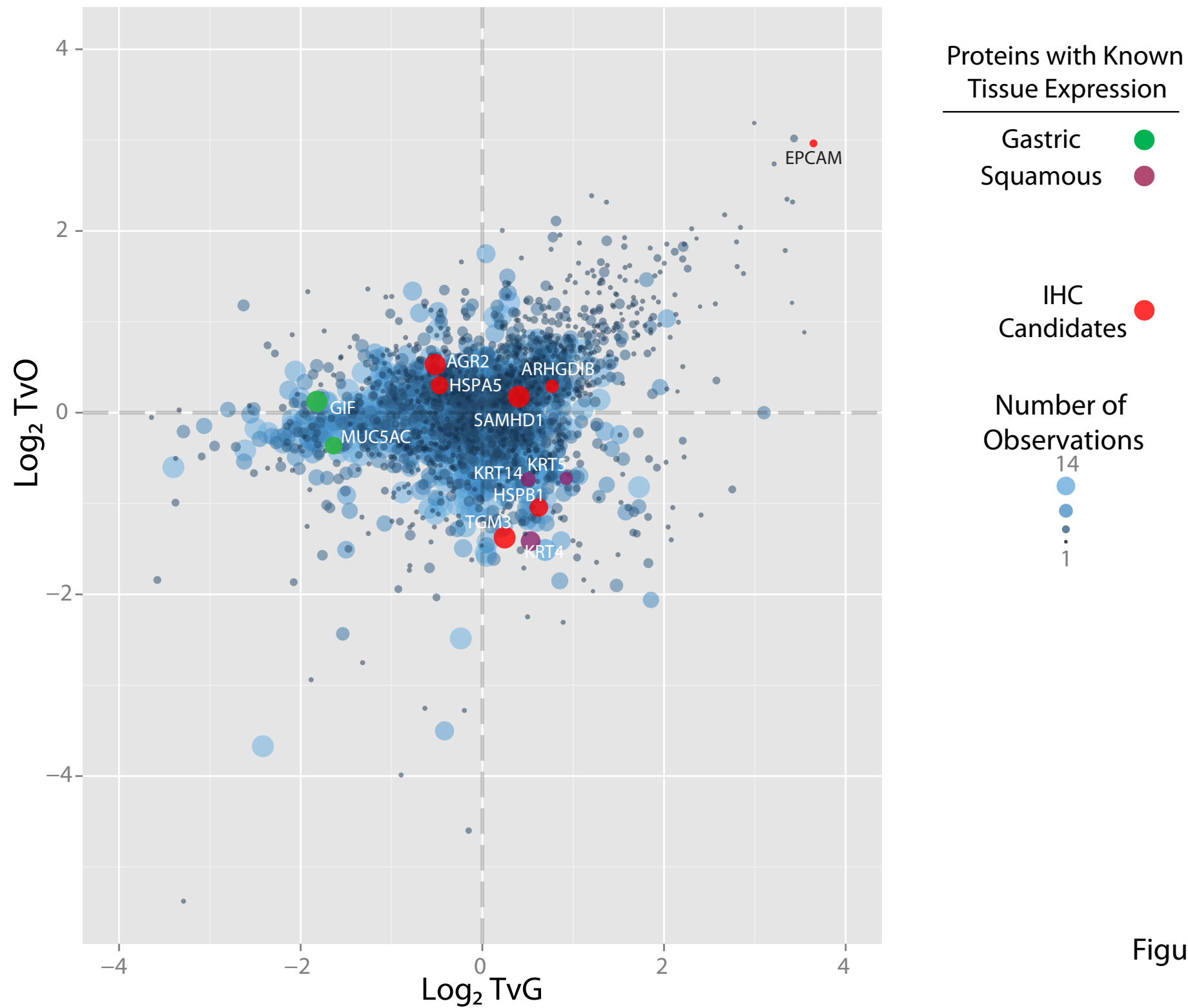


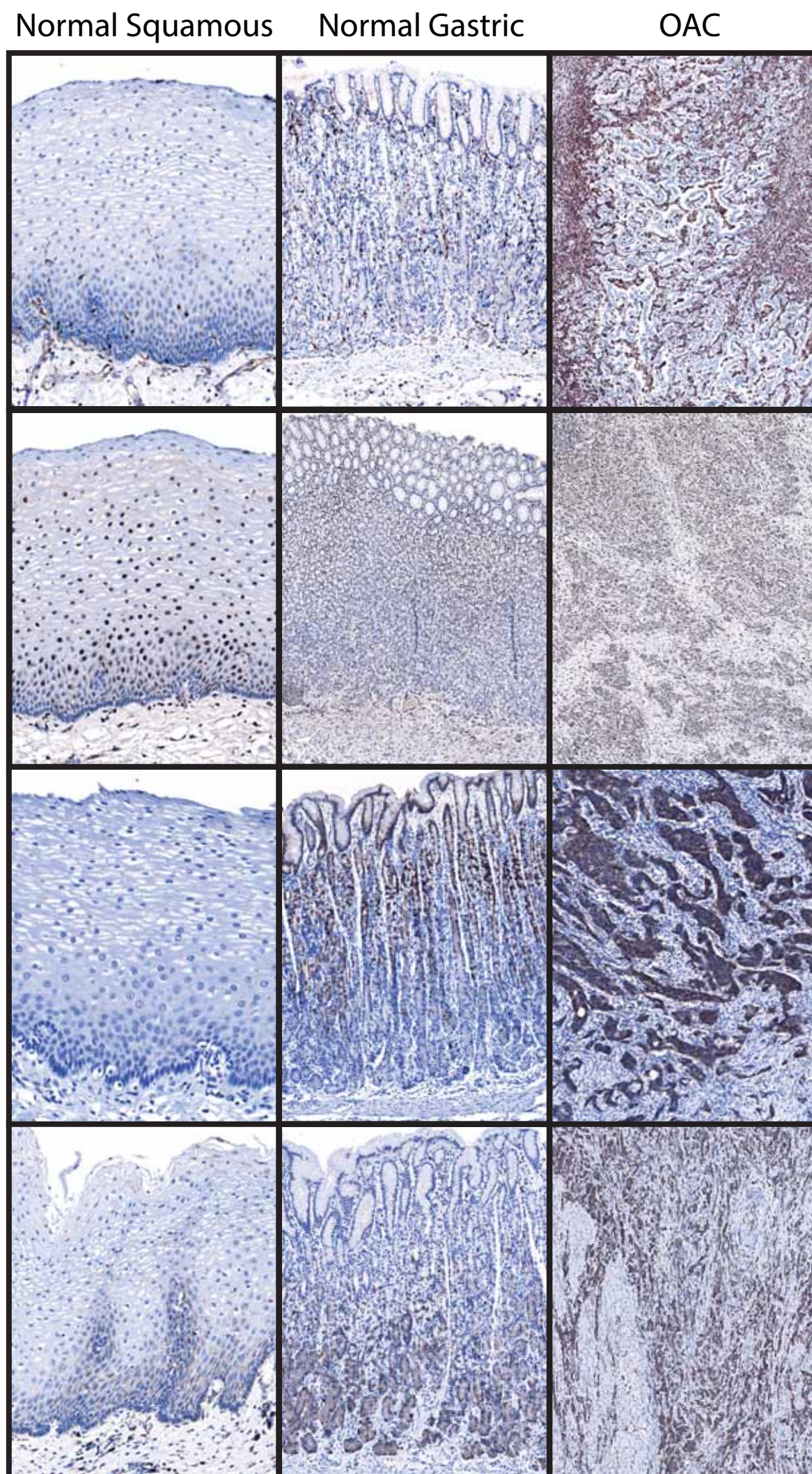
Figure 3

RHGDIB

AMHD1

AGR2

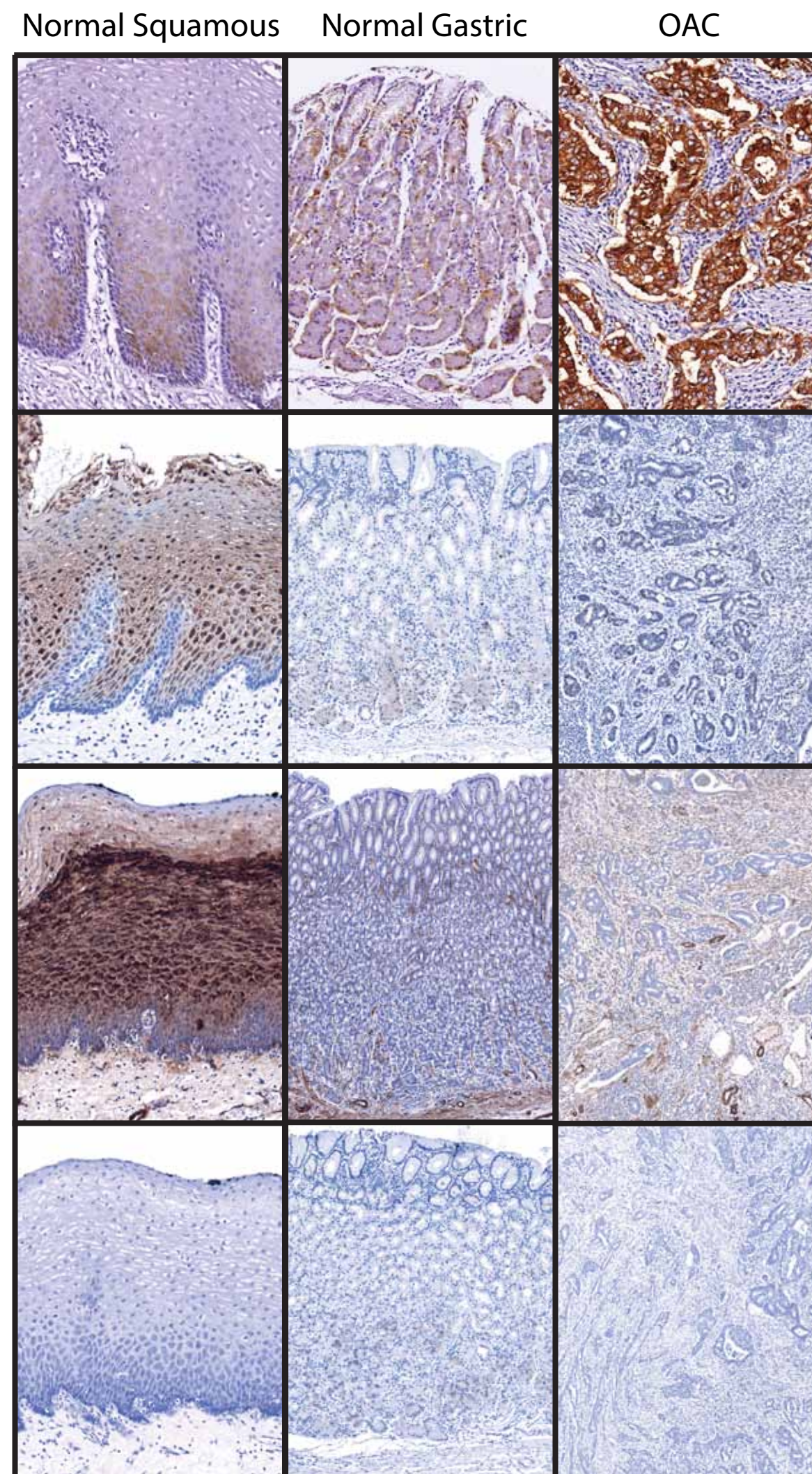
HSPA5



EPCAM

TGM3

HSPB1

No Primary
Antibody

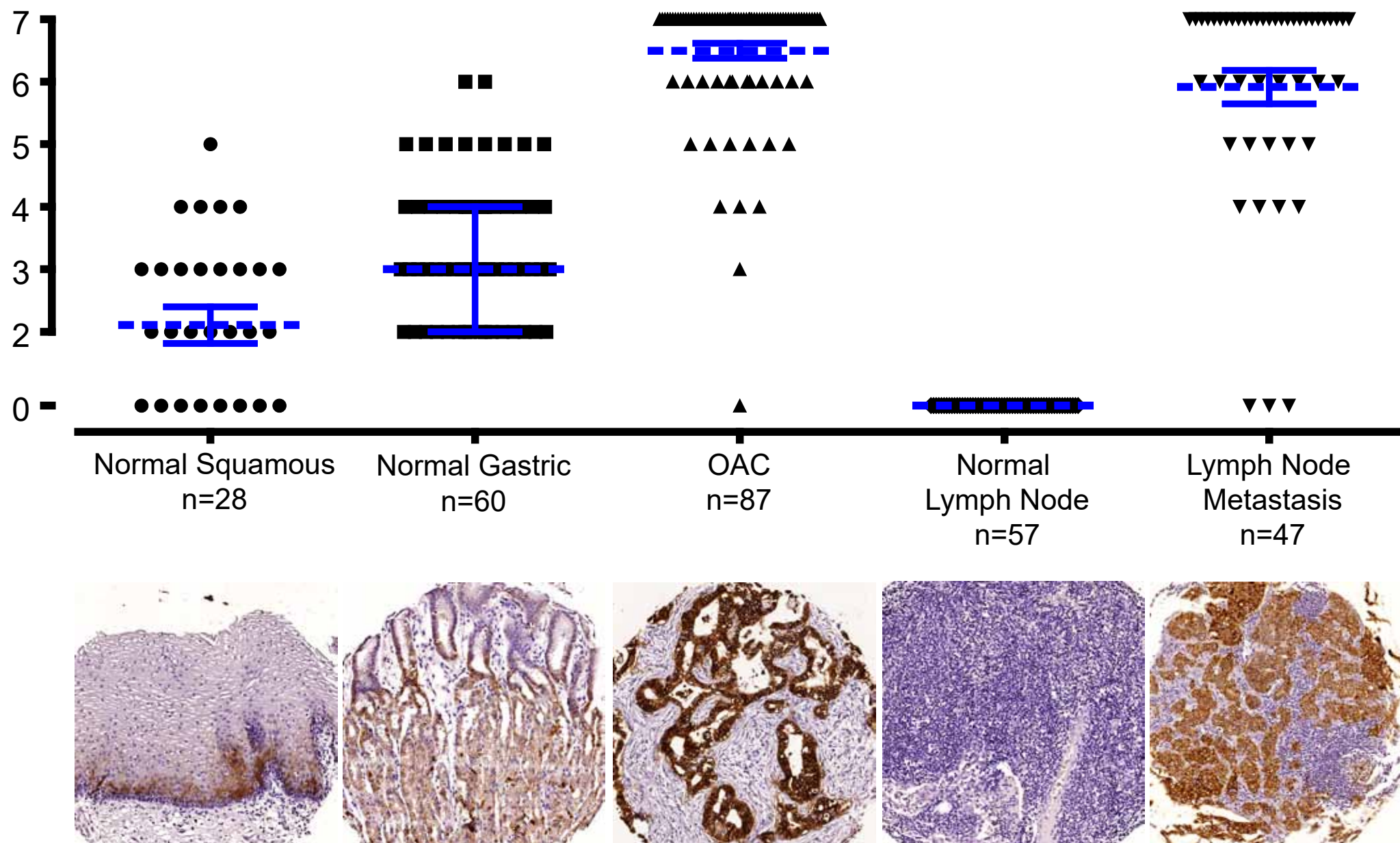


Figure 5

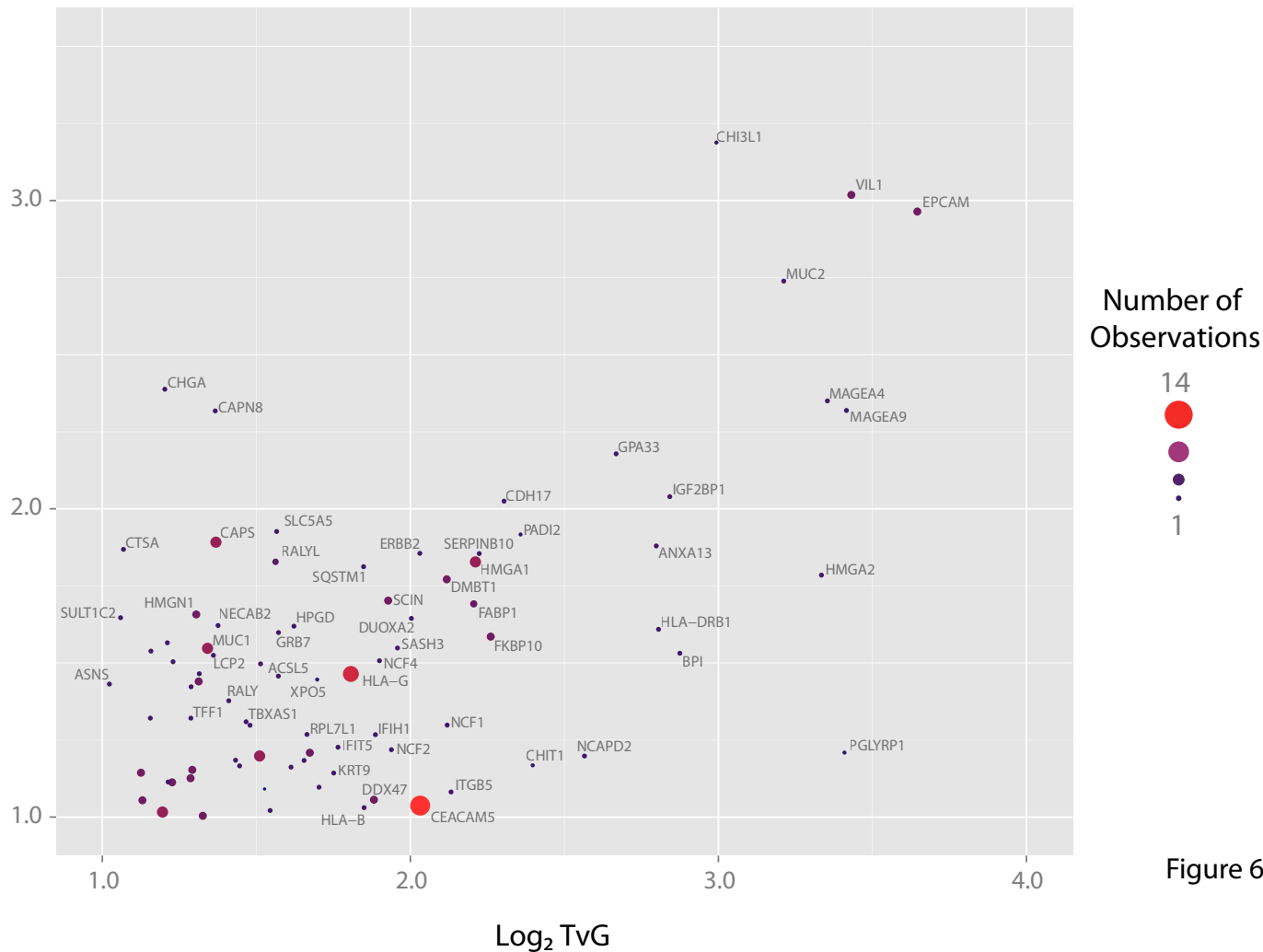


Figure 6